When AI Eats the Healthcare World - Is Trusting AI Fed, or Earned?

Syaheerah Lebai Lutfi, Adhari Al Zaabi, Geshina Ayu Mat Saat

Submitted to: Journal of Medical Internet Research
on: March 23, 2025

Table of Contents

When AI Eats the Healthcare World - Is Trusting AI Fed, or Earned?

Syaheerah Lebai Lutfi[1, 2] PhD; Adhari Al Zaabi[1] MD, PhD; Geshina Ayu Mat Saat[2] PhD

[1] Sultan Qaboos University Al-Seeb, Oman OM
[2] Universiti Sains Malaysia Penang MY

Abstract

Background: Perception-based studies are susceptible to bias introduced through the design of the instruments used. We demonstrate the need to shift from perception-based to usage-based trust evaluation, emphasizing that trust must be earned through demonstrated reliability rather than assumed from pre-adoption surveys. Our findings suggest that successful AI implementation requires a proactive approach that addresses the complex interplay of human, technical, and organizational factors, grounded in real-world usage data rather than theoretical, perception-driven acceptance measures.

Objective: To examine the disconnect between pre-adoption expectations and post-implementation realities of AI in healthcare systems.

Methods: We assessed the key perceptive-driven models, namely the Unified Theory of Acceptance and Use of Technology (UTAUT), the Technology Acceptance Model (TAM), and Diffusion of Innovation (DOI) with regards to pre-adoption of AI in healthcare. We then matched the expectations from studies using these pre-adoption models and the real results using post-usage evidences.

Results: Through empirical and anecdotal evidence, this paper demonstrates a disconnect between perception-driven technology adoption frameworks and real-world usage, focusing on the human factors that influence AI adoption and shortcomings in current perception-focused trust research.

Conclusions: Real-world usage demonstrates that hype and pre-adoption expectations fall short, and underly the reluctance or resistance of healthcare providers to fully adopt AI.

(JMIR Preprints 23/03/2025:74371)
DOI: https://doi.org/10.2196/preprints.74371

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

When AI Eats the Healthcare World - Is Trusting AI Fed, or Earned?

Syaheerah Lebai Lutfi, PhD[1, 2], Adhari Al Zaabi, MD, PhD[1], and Geshina Ayu Mat Saat, PhD[2]

[1]Sultan Qaboos University, Sutanate of Oman

[2]Universiti Sains Malaysia, Malaysia

Abstract

Objective: To examine the disconnect between pre-adoption expectations and post-implementation realities of AI in healthcare systems.
Materials and Methods: We assessed the key perceptive-driven models, namely the Unified Theory of Acceptance and Use of Technology (UTAUT), the Technology Acceptance Model (TAM), and Diffusion of Innovation (DOI) with regards to pre-adoption of AI in healthcare. We then matched the expectations from studies using these pre-adoption models and the real results using post-usage evidences.

Results: Through empirical and anecdotal evidence, this paper demonstrates a disconnect between perception-driven technology adoption frameworks and real-world usage, focusing on the human factors that influence AI adoption and shortcomings in current perception-focused trust research.
Discussion: Perception-based studies are susceptible to bias introduced through the design of the instruments used. We demonstrate the need to shift from perception-based to usage-based trust evaluation, emphasizing that trust must be earned through demonstrated reliability rather than assumed from pre-adoption surveys. Our findings suggest that successful AI implementation requires a proactive approach that addresses the complex interplay of human, technical, and organizational factors, grounded in real-world usage data rather than theoretical, perception-driven acceptance measures.

Conclusions: Real-world usage demonstrates that hype and pre-adoption expectations fall short, and underly the reluctance or resistance of healthcare providers to fully adopt AI.

Keywords: AI Trust, Perceived Trust, AI Adoption Framework, Perception-driven Framework, Usage-driven Framework

Introduction

The integration of AI into healthcare systems is increasingly common, with regulatory bodies like the U.S. Food and

Drug Administration approving several AI tools for clinical use, while others remain under evaluation. Despite the promising potential of AI in healthcare, its adoption has not kept pace with the growing body of literature and the strategic initiatives of various countries. The gap identified in Venkatesh[53], over two decades ago, which highlights three main challenges—technological barriers, human-centric factors, and regulatory constraints—remains highly relevant today, especially in the context of AI adoption in healthcare. Among these, human-centric factors—such as user acceptance, trust in AI systems, and perceived ease of use—are particularly significant. Even highly advanced technologies may struggle to achieve real-world adoption due to these factors, especially in high-stakes environments like healthcare, emphasizing the need to address them. A deeper understanding of the barriers and enablers to effective AI integration can provide valuable insights for policymakers, healthcare providers, and technology developers in constructing strategies to tackle these challenges.

Given these complexities, a key consideration is whether healthcare providers genuinely trust AI decision-making processes, or if the excitement around AI is largely *overhyped*. Evidence suggests that transparent, consistent, and predictable AI

systems are more likely to gain user trust, as demonstrated in[17,34,39,44,58,60]. Building trust is particularly crucial for healthcare providers, who are primary users of AI systems. Asan et al. [9] found that 78% of clinicians prioritize clear explanations of AI decisions, while Nundy and Montgomery [41] identified trust as a key factor driving adoption, often influenced by word-of-mouth. Moreover, Chang et al. [15] highlighted explainability as the most important determinant of trust in AI systems. This notion is not new—Lee and See [36] demonstrated that consistent behavior, even if imperfect, fosters more trust than unpredictable excellence. Ghassemi et al. [21] contends that current explainability methods are unlikely to meet expectations for patient-level decision support, describing this as “a false hope.”

While the aforementioned studies emphasize that trust in AI decision-making depends heavily on explainability, the decision to adopt AI in healthcare is often guided by pre-adoption surveys. These surveys are commonly framed using adoption models like the Unified Theory of Acceptance and Use of Technology (UTAUT)[53], the Technology Acceptance Model (TAM)[18], and Diffusion of Innovation (DOI)[45]. However, while pre-adoption surveys often highlight AI’s potential benefits, post-adoption experiences may yield different insights.

This opinion article explores this disconnect by analyzing both empirical data and anecdotal evidence, with a focus on human factors influencing AI adoption.

The paper begins by exploring whether trust in AI is “fed” or “earned,” supported by relevant evidence in the section “Trusting AI is Fed, or Earned?”. “Perception-Driven AI Adoption Models” provides an overview of perception-driven AI adoption models, while “What We May Have Missed” examines the human-centered factors that influence AI adoption in healthcare. A key focus is analyzing how perceptual trust leads to gaps between pre- and post-AI adoption experiences in “Potential Reasons for Pre vs. Post-adoption Discrepancies”, highlighting biases in traditional AI adoption models. The paper concludes by proposing usage-based models to address these gaps, emphasizing the need for sustainable, equitable, and effective AI implementation in “Narrowing the Gap: Usage-centric Trust Frameworks”. This approach aims to build the trust necessary for relying on AI to improve healthcare outcomes and shift from theoretical acceptance to meaningful, evidence-

based integration.

Table 1. Utilizing TAM/UTAUT in a Healthcare Setting

TAM (Davis, 1989)	UTAUT (Venkatesh, 2003)	Healthcare Scenario
Perceived Use (PU): The extent to which an individual believes that using the technology will enhance their performance.	**Performance Expectancy (PE):** Similar to TAM's perceived usefulness, it refers to the degree to which using the technology improves job performance.	A hospital implementing AI-driven predictive analytic in patient monitoring must ensure that clinicians recognize the performance benefits.
Perceived Ease of Use (PEOU): The degree to which an individual believes that using the AI-based predictive system will be free of effort.	**Effort Expectancy (EE):** Corresponds to TAM's perceived ease of use and relates to the effort required to use the technology.	The interface of predictive system is intuitive, easy to learn, memorize, and navigate– resulting to positive user experience (UX).
	Social Influence (SI): The degree to which users perceive that important others (e.g., peers, supervisors) believe they should use the technology.	Peer recommendations and organizational directives (top-down) to adopt the system.
	Facilitating Conditions (FC): The extent to which users believe the necessary resources and support are available to use the technology.	Periodic training sessions conducted on how to use the system, and the right kinds of tools provided.

Do Healthcare Providers *Really* Trust AI?

Trusting AI is Fed, or Earned?

It is essential to recognize that much of the discourse of AI's potential in healthcare is speculative. This does not imply that experts are uninformed or intentionally misleading but rather that the narrative often hinges on a theoretical future. Conversations about AI's greatest capabilities frequently emphasize a visionary. Do Healthcare Providers *Really* Trust AI? outlook, portraying current limitations as temporary hurdles. Present-day challenges are often downplayed in favor of highlighting an anticipated future where AI achieves its full promise—-a future that, for now, remains largely aspirational. At this juncture, healthcare providers are supposed to accept this situation at face value, without verification or concrete evidence.

In the healthcare sector, the adoption of AI systems often walks a fine line between speculative assurances and evidence-based decision-making. Developers and advocates frequently claim that current limitations—such as hallucinations, unreliable interfaces, and unpredictable behavior[56]—will resolve as the technology matures. While this optimism may drive some adoption decisions, it risks overshadowing critical human factors, including limitations in healthcare practitioner knowledge and technical skills[8,19,37], the steep learning curve for users[1,2] that could cause cognitive overload leading to burnout[10] and the inherent subjectivity in medical case management[3]. These factors, in turn, influence decisions related to organization-wide adoption, case management practices, and system security.

The prevailing narrative, however, suggests that adoption decisions are typically based on comprehensive pre-adoption qualitative assessments. These evaluations aim to verify whether an AI system aligns with both clinical and non-clinical needs, and determine its suitability for implementation. But do they truly accomplish this? Questions remain about their ability to fully address key challenges such as cost-effectiveness, long-term maintenance, dependency on technology and machinery, disruptions from power shortages or system failures, human error,

competency training, and adherence to ethical standards—including access, accountability, confidentiality, and patient data sharing.

The two key qualitative perception-driven models that underpin our understanding of human perception in technology adoption are the Technology Acceptance Model (TAM)[18] and the Unified Theory of Acceptance and Use of Technology (UTAUT)[53]—and their relevance to healthcare settings are discussed in the next section.

Perception-Driven AI Adoption Models

While these models are intended to assess organizational readiness for technology adoption, they often fail to capture resistance to change driven by the sentiment of why change something that already works? This limitation arises because these frameworks are largely perception-driven, focusing on initial beliefs rather than actual outcomes. Additionally, the design of instruments based on these models can be manipulated through biased elements to produce outcomes that align with organizational agendas—such as justifying funding allocations or demonstrating public acceptance, regardless of genuine readiness. In the section "What We May Have Missed", further evidence demonstrates how pre-adoption perceptions often fail to align with real-world experiences, revealing the true trust in AI only after practical implementation.

- **The Technology Acceptance Model (TAM)**, developed by Davis in 1989[18], is a widely used framework for studying technology adoption. It identifies two key factors influencing whether a user accepts and uses new technology: Perceived Usefulness (PU), which reflects the belief that the technology will enhance performance, and Perceived Ease of Use (PEOU), which refers to the belief that using the technology will be effortless. According to TAM, these factors shape users' attitudes toward the technology, which subsequently affect their *behavioral intention* to use it and later, their actual usage behavior. In the context of AI in healthcare, perceived usefulness might pertain to benefits such as improved diagnostic accuracy or faster decision-making, while perceived ease of use could depend on seamless integration into existing workflows and intuitive system interface design. For instance, radiologists are more likely to adopt AI-powered diagnostic tools if they perceive these tools as significantly enhancing diagnostic speed and reducing errors (PU) and if the systems are intuitive and require minimal training (PEOU).

- **Unified Theory of Acceptance and Use of Technology (UTAUT)**, proposed by Venkatesh et al. [53], incorporates a number of earlier models, majorly the Technology Acceptance Model (TAM). It pinpoints four crucial elements affecting adoption: Effort Expectancy (EE), which reflects the effort required to use the technology and correlates with TAM's perceived ease of use; Performance Expectancy (PE), which relates to the degree to which the technology improves job performance and aligns with TAM's perceived usefulness; and Social Influence (SI), which takes into account the degree to which users believe that significant individuals or groups (e.g., peers, supervisors) think they should adopt the technology; as well as Facilitating Conditions (FC), which describe the user's belief that the required tools and assistance are accessible.

- **Example of Relevance of the TAM/UTAUT Components in Healthcare (see Table 1)** In healthcare, the adoption of AI tools is primarily shaped by perceptual factors, as outlined in the TAM and the Unified Theory of UTAUT. Clinicians often base their decisions to adopt AI

technologies on social influence, such as peer recommendations, evidence-based practices elsewhere and organizational directives (top-down). The successful integration of AI also depends on facilitating conditions, including access to training programs and IT support. For example, when a hospital introduces AI-driven predictive analytics for patient monitoring, it is crucial that clinicians perceive the technology's performance benefits (Performance Expectancy, PE), find it easy to use (Effort Expectancy, EE), receive endorsements from leadership or peers (Social Influence, SI), and have adequate training and infrastructure (Facilitating Conditions, FC) to support its implementation. In healthcare, providers focus less on the technical details of AI and more on the value it brings to practice, particularly in improving patient safety and care.[12].

What We May Have Missed

In reality, discussions regarding the divergence between users' expectations of AI use and the actual use of AI in healthcare and decisions to continue or replace AI systems in healthcare; are often underrepresented in academic literature. Healthcare academia often under-emphasizes "contrary" evidence highlighting challenges in the real-world use of AI in healthcare. Instead, these issues are predominantly discussed in blog posts and newsletters authored by AI commentators[14,56] and industry observers. These writers frequently argue that many perspectives on AI in healthcare are shaped by the broader discourse surrounding AI *hype*, which is often driven by a Fear of Missing Out (FOMO) narrative. However, a small number of academic reviews do address this divergence.
Section present evidence from multiple sources, both anecdotal and empirical, demonstrating AI healthcare adoption discrepancies - illustrating that What You See Is *Not* What You Get (WYSI*N*WYG). These findings challenge the often optimistic pre-adoption projections with post-implementation realities.

Perceived vs. Actual Trust

Over the past five years, numerous systematic literature reviews (SLRs) have highlighted clinicians' perceptions of AI

decision aids as valuable tools[26,33,43,48,59]. For instance, Hassan et al. [26] and Younis et al. [59] showed AI's potential as a promising shared decision-making (SDM) tool, while Perivolaris et al. [43] emphasized its perceived contribution to improving the quality of clinician-patient interactions. Furthermore, several SLRs have reported AI achieving either superior[32] or comparable[48] performance in offline modeling—where non-real-time data from trials are analyzed—or in controlled laboratory and simulated environments.
The term perceive is particularly significant here, as it circumscribes these findings within the domain of clinicians' expectations and theoretical beliefs, rather than empirical implementation outcomes. While extant research predominantly examines the *potential utility* of these systems, empirical investigations of clinical implementations reveal substantial operational challenges. For instance, Johri et al. [30] document unrealistic doctor-patient interactions, while Ankolekar et al. [6], Mooghali et al. [40] highlight how insufficient validation processes impede trust formation. Although Shared Decision Making (SDM) frameworks demonstrate promise in facilitating theoretical discourse, their efficacy diminishes markedly during practical clinical encounters. Furthermore, Botha et al. [13] present compelling evidence that AI models, despite demonstrating impressive predictive accuracy in controlled environments, often exhibit performance approximating random chance when deployed in clinical and emergency settings.

Further substantiating these concerns, Steerling et al. [50]'s comprehensive scoping review identifies a critical lacuna in contemporary AI trust research within healthcare settings. While the literature

extensively examines trust through multiple theoretical lenses—encompassing technological capabilities, interpersonal dynamics, and mediating factors—there exists a notable absence of empirical investigation into trust within the context of *actual system usage*. Their analysis reveals a predominant focus on individual perceptions of AI capabilities and interpersonal trust dynamics, rather than examining the evolution and maintenance of trust during practical system implementation and utilization. This methodological gap becomes particularly salient that existing research primarily interrogates technological and individual characteristics, while broader contextual factors and real-world usage patterns remain largely unexamined, or simply unreported. Such findings suggest a critical need for more comprehensive, usage-focused trust research in healthcare AI implementation.

Potential Reasons for Pre vs. Post-adoption Discrepancies

In the healthcare domain, trust, explainability, and demonstrable benefits are often cited as critical factors for the successful integration of AI systems. Ideally, structured evaluations should ensure that adoption decisions are grounded in real-world feasibility, free from the influence of biased pre-adoption information, stakeholder vested interest driven initiatives or overly optimistic projections about future capabilities.

At this juncture, while structured evaluations should ideally ground adoption decisions in real-world feasibility, questions remain about potential biases. As Warzel [56] sardonically observes, Judge us *not* by what you see, but by what we (the AI developers) imagine. This highlights the gap between promotional narratives and practical realities. Figure 1 illustrates the

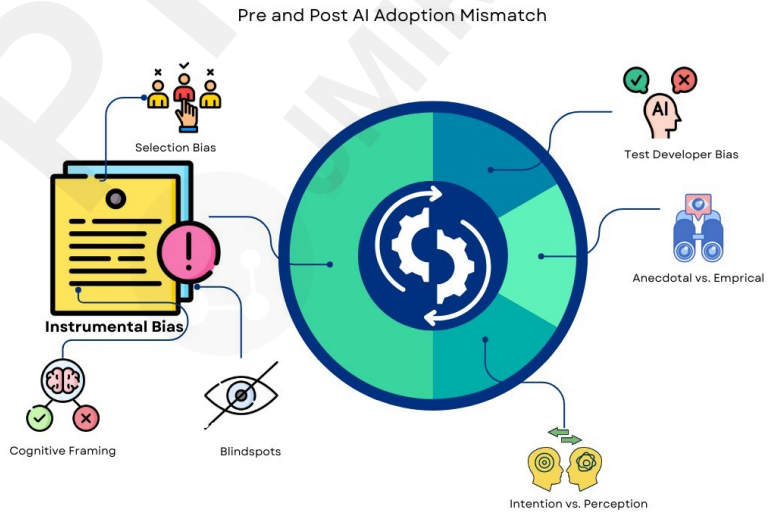


Figure 1. Potential Reasons for the Gaps between Pre and Post AI Adoption Outcomes

potential causes of mismatches between pre-adoption expectations and post-adoption realities of AI in healthcare, followed by detailed explanations.

Instrumental Bias in Technology Adoption Research

• Selection Bias in Survey Construction

Instrumental bias in technology adoption research occurs when measurement tools actively shape the knowledge being produced, rather than serving as neutral instruments for observation. Bias refers to a systematic distortion of the truth, which can occur at any stage of a study—whether in its design, execution, or analysis. In healthcare surveys, particularly those examining technology adoption, selection bias can significantly affect the validity and generalizability of the results. Recognizing and addressing selection bias is crucial for translating evidence into effective clinical practice. In their seminal work, Hegedus and Mood [29] comprehensively delineate more than 40 distinct categories of selection bias see Tables 1 & 2 in[29]. Despite its publication over a decade ago, this taxonomical framework remains foundational to contemporary discussions of methodological bias in research design.

Narrowing down to technology adoption, selection bias may involve choosing individuals who are tech-savvy or unintentionally (or intentionally) focusing on groups with higher education levels or younger individuals, who are generally more comfortable with technology[51]. Similarly, selectively including or excluding participants based on factors like geographic location, socioeconomic status, or a predisposition to provide positive feedback on successful adoption stories—while overlooking challenges or failures can create an overly optimistic portrayal of technology adoption. This often leads to nonadoption or abandonment of the technology, despite organizational or policy

support[23,24,49].

• Cognitive Framing

Cognitive framing of AI adoption tools refers to the way in which respondents cognitively interpret and process survey questions, influenced by the way these questions are structured and presented. Contextual information has a significant impact on response formation. Responses to surveys on the introduction of AI can be strongly influenced by the immediate context in which the questions are asked. For example, questions about perceptions of AI capabilities may elicit different responses depending on whether they are preceded by scenarios that emphasise positive outcomes (e.g. improved diagnostic accuracy) or negative outcomes (e.g. concerns about ethical risks). The order in which questions are asked can also influence respondents' answers. For example, if respondents are asked about ChatGPT first and later asked to name the most effective generative AI system, their answers may be biassed towards ChatGPT, even though other generative AI systems (such as GPT-4, Bard, or other industry-specific tools) that do not readily come to mind are relevant.

Similarly, questions that tap into the respondent’s individual trait rather than directly asking about their views on or adoption of AI technologies in healthcare, lead to bias response (e.g., How confident are you in your ability to adapt System ABC into your professional practice?) Such questions can unintentionally introduce bias if they are used to gauge AI adoption indirectly, as the answers might be more reflective of the individual’s self-perception and personal characteristics (like confidence, comfort with technology, or past experiences) rather than their actual stance on AI
itself.

Surveys that rely heavily on jargon or contain an excessive number of questions can lead to response fatigue if respondents prioritise quick completion over thoughtful and accurate answers. In addition, the order of questions can lead to bias. For example, starting with a series of sensitive or emotionally charged questions can affect the way respondents react to subsequent questions. Taken together, these factors increase cognitive load, which can affect the quality and reliability of the data collected.

• Blindspots in Mental Health Instruments

In the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th Edition)[5], NOS (Not Otherwise Specified) serves as a category for cases where symptoms do not fully meet the criteria for a specific mental disorder but still represent significant clinical concerns. More recently, NOS has been replaced by the classifications *Other Specified Disorder* (OSO) or *Unspecified Disorder* (UD), depending on the context of the diagnosis.
Similarly, in the *International Classification of Diseases (ICD-11)*[57], NOS also refers to Not Otherwise Specified and functions as a subcategory for general diagnoses where a more specific classification cannot be assigned. This designation is employed when a clinician cannot fit the presenting symptoms into an established diagnostic category. In contrast, Not Elsewhere Classified (NEC) is used when a detailed diagnosis is available but the classification codes lack the specificity to accommodate it.

The use of NOS, OSO, UD, and NEC could represent a *blindspot* for AI systems in mental health instruments. While these categories rely on nuanced clinical judgment to address cases that fall outside well-defined diagnostic criteria, AI systems may lack the capacity to replicate such judgment effectively. These blindspots emerge because:

1. AI development may overlook scenarios due to human limitations such as forgetfulness, incomplete knowledge, lack of empirical evidence for specific conditions, or the complexity of co-occurring symptoms.
2. The system’s design is often more reflective of the perspectives and biases of developers than the lived experiences of healthcare professionals and patients.

This reliance on general or nonspecific categories in mental health instruments can undermine the effectiveness of AI in supporting clinical decision-making. It risks increasing the frequency of NOS, OSO, UD, and NEC categorizations, thereby limiting the system’s capacity to provide objective insights or guide clinicians toward optimal patient care. Furthermore, these categorizations create cascading challenges in healthcare delivery–particularly in emergency settings and referencing across different systems and healthcare institutions. They complicate insurance coverage and home care planning, especially in terms of financial accountability and projecting wellness milestones.
Without addressing this *blindspot*, follow-up appointments are likely to result in frustration for patients, families, and providers, as treatment strategies will remain limited to maintaining the

current state of health or preventing further decline, rather than achieving meaningful improvement. Instrument development studies that involve clinicians with extensive experience in managing provisional or uncertain diagnoses are essential to mitigate these issues and enhance the applicability of AI in mental health care.

Misalignment Between Test Intent and User Perception

A particular instrument may be created where the intended purpose of a test diverges from the perceptions or expectations of its users. For instance, a diagnostic AI tool designed for depression screening might be misunderstood by users as capable of predicting suicidality, leading to responses not targeted for the intended questions.

Test Developer Bias

Test developer bias extends beyond technical design flaws to encompass the lived experiences of test-takers and the broader implications of application. Many assessments, particularly those measuring motives in criminal behavior—such as attempted murder—fail to account for the nuanced realities of both perpetrators and victims. These tests often operate under rigid frameworks that overlook the complexity of individual experiences, leading to potential misinterpretations and unfair conclusions.

A test-taker's unique emotional, educational, and experiential background significantly influences their engagement with survey items. Factors such as pain, illness, disease, personality traits, and even the physical environment in which a test is taken can introduce unintended bias. Additionally, elements like language proficiency, educational background, gender, culture, and religious beliefs shape both responses and result interpretations. For instance, cultural norms play a crucial role in how individuals perceive and interact with healthcare AI technologies, yet many standardized tests fail to account for these differences.

Anecdotal Evidence Beyond Empiricism

Decision-making related to the adoption of artificial intelligence (AI) in healthcare often hinges on both anecdotal and empirical evidence, each providing distinct insights and challenges.

Anecdotal evidence in healthcare stems from personal experiences, testimonials, and case studies of medical practitioners and patients. This type of evidence offers valuable insights into the practical applications and immediate impacts of AI technologies in real-world settings[22,25,47]. For instance, triage personnel may report cases where AI-assisted diagnostics led to early detection of rare diseases, facilitated risk stratification, or identified conditions requiring immediate intervention. Despite its utility, anecdotal evidence is often

constrained by a lack of generalizability, statistical robustness, and potential biases stemming from overlapping symptoms or institutional variations in standard operating procedures.

Empirical evidence, derived from systematic research methods such as randomized controlled trials, observational studies, and systematic reviews, provides a higher level of rigor and reproducibility. Such evidence forms a stronger basis for informed decision-making[4,32,35,54]. For example, systematic reviews might analyze multiple studies to assess the effectiveness of AI in expediting diagnoses and enhancing patient outcomes. However, empirical findings may overlook nuanced individual experiences, particularly in emergency or complex cases. Consequently, a balanced approach incorporating both anecdotal and empirical evidence is critical for the effective integration of AI in healthcare.
The following use cases illustrate how anecdotal evidence should not be disregarded or treated as yet another blind spot.

Anecdotal insights, while lacking the statistical rigor of empirical research, provide valuable contextual understanding and practical nuances that are often missed in large-scale studies. These cases demonstrate the critical role anecdotal evidence can play in identifying gaps, refining implementation strategies, and complementing empirical findings:

- **Recruitment Profiling System (SPIER):** The Recruitment Profiling System (SPIER) is an online platform employing psychological testing as an initial screening tool in the recruitment process for new officers. Developed collaboratively by various agencies under the Malaysian Ministry of Home Affairs[38], SPIER is currently utilized by 12 agencies.

  According to personal communication, SPIER 1.0 faced several challenges[7] (13 January 2025). Due to its success in screening new officers, other agencies expressed interest in adopting its assessments for their recruitment processes. However, some individuals exploited SPIER 1.0 by accessing it under false pretenses to replicate its domains and items, leading to unethical practices such as creating cheat pathways, offering coaching services, or developing unauthorized assessment tools[7] personal communication (13 January 2025). SPIER 2.0 has since implemented enhanced safeguards and improved the reliability of its assessments.

- **Traditional and Electronic Record-Keeping:** At Universiti Sains Malaysia Specialist Hospital, both traditional (paper-based) and electronic methods are employed to record and maintain patient information. Each method has its vulnerabilities. Traditional records are susceptible to storage limitations, misplacement, and ink degradation[28] personal communication (14 January 2025). On the other hand, electronic records may become inaccessible during power outages or suffer delays due to poor internet connectivity[27] personal communication (15 January 2025). As highlighted by healthcare providers, these dual systems are essential to ensure continuity and reliability under varying conditions[27,28].

  Systems like SPIER and electronic records highlight that trust in AI is often given without scrutiny, despite risks like misuse and technical failures. True trust must be earned through consistent reliability and ethical safeguards, not assumed.

Discussion

Narrowing the Gap: Usage-centric Trust Frameworks

As discussed above, perception-driven approaches to evaluating AI trust are biased and lack objective measures based on actual usage. This paper proposes a shift toward usage-centric trust frameworks (also known as the human-centric model), which focus on trust earned through real-world reliability. We recommend developing holistic instruments based on these frameworks, shifting the focus from subjective opinions and hypothetical expectations to empirical evidence grounded in actual implementation and use.

Design Thinking for User Experience (Design Thinking UX)

Healthcare User Experience (UX) research is more prevalent in industry reports and private sector evaluations than in academic literature, yet academia should adopt similar frameworks to enhance evidence-based understanding of AI adoption and user trust. As researcher and nurse from Penn Nursing, Walker [55], explains, there are huge gaps in our ability to translate our innovation out into the real world. Design Thinking UX, which is increasingly applied AI-based healthcare product development, offers a practical approach to inform AI adoption. This aligns with a ”realist framework,” which rejects one-size-fits-all models[42], recognizing that outcomes vary depending on the context and population. By considering the specific conditions and mechanisms influencing adoption, this framework offers deeper insights that can help developers and policymakers ensure effective AI integration in healthcare.

A major challenge for clinicians and patients when using healthcare applications is the difficulty in navigating the system efficiently and effectively, which can also pose risks to patient safety. UX design directly impacts user adoption, task completion time, error reduction (enhancing patient safety), fewer support requests, and overall user satisfaction[16].

A key advantage of Design Thinking UX is its emphasis on human-centered, *iterative* approach. It is not just about theoretical readiness but about *continuously* improving and adjusting the system based on user feedback and real-world usage. This iterative process allows for a dynamic response to issues that emerge during the adoption of AI systems, addressing both user experience and organizational needs as they evolve. In healthcare, where contexts and workflows are highly variable, Design Thinking allows for adaptive solutions that can better cater to specific needs at different, five non-linear stages of AI adoption. These are *Empathize (Learn), Define, Ideate, Prototype, Test* (see Figure 2).

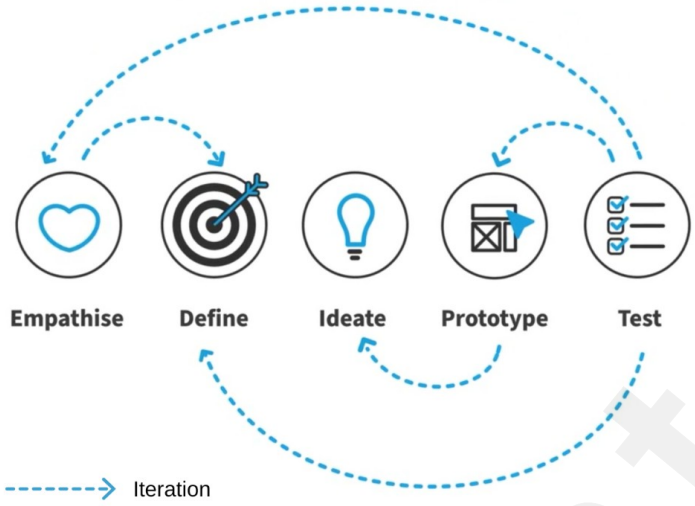


Figure 2. The five stages of iterative, non-linear Design Thinking Framework (Adopted from[20])

.

Extrapolating from the work of Seneviratne et al. [46], designing AI-driven systems to fit outdated data under controlled conditions, rather than addressing the complexities of contemporary clinical settings, has led to challenges in the adoption of machine learning (ML) models in healthcare and case management. Essentially, this work points out a disconnect between how AI models are trained and the real-world conditions they are expected to operate in. To mitigate this issue, Seneviratne et al. [46] proposed a practical care management toolkit based on user-centered design. This toolkit includes: (1) addressing solvable pain points, (2) leveraging the unique value of ML (e.g., automation and augmentation), (3) creating an actionable pathway, and (4) defining the model's reward function. While the results of this study are promising, it requires periodic case management reviews by relevant stakeholders, which can be time-consuming due to the wide range of cases and the involvement of multiple stakeholders. Additionally, the toolkit faces challenges in responding swiftly to the dynamic nature of health symptoms and treatment progress. Nevertheless, the focus on user-centered design, iterative refinement based on user feedback, and addressing real-world pain points mirrors key aspects of the Design Thinking methodology.

Explanation User Interface (XUI)

More recently, healthcare informatics researchers in academia have advocated for approaches that build trust through transparency and informed knowledge, ensuring that healthcare providers understand a system's purpose, decision-making processes, and performance in *context*. This reflects

the iterative and human-centered nature of Design Thinking, which is evident in recent work emphasizing the use of mock-ups that actively involve healthcare practitioners as users in evaluating AI applications in healthcare throughout the development process—not just at the beginning or end, but in a continuous,

iterative manner [11,31,52].

Healthcare stakeholders, whether providers or patients, interact with the user interface as their first point of engagement with an AI system. To ensure effective usability, researchers use partially functional prototypes or mock-ups to evaluate real-world navigation and interaction. Beyond usability, the system should also provide explanations that help users understand its output and facilitate appropriate use. This explanatory component, integrated within the user interface, is known as the explanation user interface (XUI)[31]. By incorporating both usability testing and explainability, this approach offers a more rigorous and empirical evaluation compared to relying solely on surveys or questionnaires, which often capture perceptions based on abstract expectations rather than direct user experience.

In this context, Jung et al. [31] performed a scoping review, yielding 64 recommendations structured into five branch question-based mind maps for the user-centered design of explanation user interfaces (XUIs) in artificial intelligence (AI)-driven clinical decision support systems (CDSS) in healthcare. These five-branch mind maps addressed:

- Branch 1: What information to provide in XUIs?
- Branch 2: How to design XUIs?
- Branch 3: How to evaluate XUIs?
- Branch 4: How to configure the design process of XUIs?
- Branch 5: What to keep in mind when designing XUIs?

A comprehensive mind maps for all the branches is available in their paper. Among others, an interesting point raised was the transferability of recommendations made across different healthcare institutions, which require further testing and review. Nevertheless, this progression underscores the growing recognition of human-centered, iterative methodologies in AI adoption for healthcare.

Conclusion

AI implementation in healthcare must be proactive rather than reactive, requiring a comprehensive consideration of human and technical factors before deployment. Trust is central to this process, influencing initial acceptance and long-term integration. This paper highlights the critical gap between pre-adoption expectations and real-world implementation, particularly in how trust in AI systems is formed and sustained.

Traditional technology acceptance models, such as TAM and UTAUT, while foundational, often fall short in capturing the complexities of AI in healthcare, as they rely heavily on perceived usefulness rather than practical readiness. This perceptual bias can lead to inflated expectations, often fueled by hype rather than alignment with clinical realities. As real-world constraints—such as workflow disruptions, infrastructure limitations, and inadequate training—become apparent, enthusiasm alone

proves insufficient for sustained AI adoption.
Trust in AI systems remains a challenge, as misalignment between developers' designs and healthcare providers' practical requirements frequently leads to implementation issues. Real-world workflows, infrastructure limitations, and insufficient training exacerbate this gap. Enthusiasm, particularly among the younger, tech-savvy population, may stem from overhyped expectations or limited understanding of AI's constraints. However, this optimism does not guarantee sustained acceptance, as post-implementation experiences remain under-explored. Both anecdotal and state-of-the art evidences presented highlight the gap between initial excitement and the practical realities of AI integration.

To bridge these gaps, healthcare AI must transition from perception-driven models to *usage-focused* trust frameworks. These require ongoing validation through real-world application, emphasizing demonstrated reliability, usability, and evolving trust over time. The increasing integration of XUI through Design Thinking principles, whether explicitly acknowledged or not, reinforces the importance of iterative, user-centered development. By prioritizing transparency, adaptability, and human-centered evaluation, AI can be effectively integrated into clinical workflows, fostering sustained trust and delivering meaningful improvements in provider experience and patient care.

References

1. T. Abrahams, O. Farayola, S. Kaggwa, P. Uwaoma, A. Hassan, and S. Dawodu, "Cybersecurity awareness and education programs: A review of employee engagement and accountability," *Computer Science & IT Research Journal*, vol. 5, no. 1, pp. 100–119, 2024.

2. A. Aderibigbe, P. Ohenhen, N. Nwaobia, J. Gidiagba, and E. Ani, "Artificial intelligence in developing countries: Bridging the gap between potential and implementation," *Computer Science & IT Research Journal*, vol. 4, no. 3, pp. 185–199, 2023.

3. O. Akindote, A. Adegbite, S. Dawodu, A. Omotosho, A. Anyanwu, and C. Maduka, "Comparative review of big data analytics and gis in healthcare decision making," *World Journal of Advanced Research and Reviews*, vol. 20, no. 03, pp. 1293–1302, 2023.

4. A. Al Kuwaiti, K. Nazer, A. Al-Reedy *et al.*, "A review of the role of artificial intelligence in healthcare," *Journal of Personalized Medicine*, vol. 13, no. 6, p. 951, 2023.

5. American Psychiatric Association, Ed., *Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition, Text Revision (dsm-5-Tr)*. Washington, DC, USA: American Psychiatric Publishing, 2022.

6. A. Ankolekar, B. van der Heijden, A. Dekker *et al.*, "Clinician perspectives on clinical decision support systems in lung cancer: Implications for shared decision-making," *Health Expectations*, vol. 25, no. 4, pp. 1342–1351, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/hex.13457

7. Anonymous, "Personal communication on spier 2.0," January 13 2025, kelantan.

8. E. Anton, A. Behne, and F. Teuteberg, "The humans behind artificial intelligence–an operationalisation of ai competencies," in *Twenty-Eighth European Conference on Information Systems (ECIS2020) – A Virtual AIS Conference*, ser. Research Papers, no. 141, 2020, pp. 1–17. [Online]. Available: https://aisel.aisnet.org/ecis2020_rp/141

9. O. Asan, A. Choudhury, M. Bose, and J. Zhu, “Building Trust In Healthcare Artificial Intelligence: The FACTS Framework,” *Nature Digital Medicine*, vol. 6, no. 1, p. 82, 2023.

10. E. Asgari, J. Kaur, G. Nuredini *et al.*, “Impact of electronic health record use on cognitive load and burnout among clinicians: Narrative review,” *JMIR Med Inform*, vol. 12, p. e55499, Apr 2024. [Online]. Available: https://medinform.jmir.org/2024/1/e55499

11. J. Bichel-Findlay *et al.*, “Medinfo 2023 — the future is accessible,” in *MEDINFO 2023*. International Medical Informatics Association (IMIA) and IOS Press, 2024, this article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

12. T. Boillat, F. Nawaz, and H. Rivas, “Readiness to embrace artificial intelligence among medical doctors and students: Questionnaire-based study,” *JMIR Med Educ*, vol. 8, p. e34973, 2022.

13. N. Botha, C. Segbedzi, V. Dumahasi *et al.*, “Artificial intelligence in healthcare: A scoping review of perceived threats to patient rights and safety,” *Archives of Public Health*, vol. 82, p. 188, 2024. [Online]. Available: https://doi.org/10.1186/s13690-024-01414-1

14. M. Brandeis, “Labeling Your Ai Dependency,” Online, Oct. 2023, [Online; posted 20-June-2023]. [Online]. Available: https://medium.com/@brandeismarshall/labeling-your-ai-dependency-9828194877a3

15. A. C. Chang, M. H. Moniz, K. Patel *et al.*, “Trust in clinical decision support systems: A systematic review,” *Journal of Medical Internet Research*, vol. 24, no. 11, p. e41654, 2022, identified explainability as top factor influencing provider trust (87% of participants) and found previous experience significantly impacted trust levels.

16. L. Chapman, “Design thinking ux framework for healthcare apps,” n.d., accessed: 2025-01-20. [Online]. Available: https://www.emids.com/insights/ design-thinking-ux-framework-healthcare-apps/

17. S.-Y. Chien, K. Sycara, M. Lewis, and H. Liu, “Building Trust Through Predictable AI Behavior,” in *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems*. IFAAMAS, 2023, pp. 112–120.

18. F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

19. B. Familoni and N. Onyebuchi, “Advancements and challenges in ai integration for technical literacy: A systematic review,” *Engineering Science & Technology Journal*, vol. 5, no. 4, pp. 1415–1430, 2024.

20. I. D. Foundation, “Design thinking,” 2025, n.d. [Online]. Available: https://www.interaction-design.org/literature/topics/design-thinking

21. M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, “A false hope? maintaining trustworthiness in the face of artificial intelligence,” *The Lancet Digital Health*, vol. 3, no. 12, pp. e745–e750, 2021.

22. C. B. Goldberg, L. Adams, D. Blumenthal *et al.*, “To do no harm—and the most good—with ai in health care,” *NEJM AI*, vol. 1, no. 3, p. AIp2400036, 2024.

23. T. Greenhalgh, J. Wherton, C. Papoutsi *et al.*, “Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies,” *Journal of Medical Internet Research*, vol. 19, no. 11, p. e367, 2017.

24. T. Greenhalgh, J. Wherton, C. Papoutsi, J. Lynch, G. Hughes, and A’Court, “Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the nasss framework,” *BMC Medicine*, vol. 16, pp. 1–15, 2018.

25. V. Harish, F. Morgado, A. D. Stern, and S. Das, “Artificial intelligence and clinical decision making: The new nature of medical uncertainty,” *Academic Medicine*, vol. 96, no. 1, pp. 31–36, 2021.

26. N. Hassan, R. Slight, K. Bimpong *et al.*, "Clinicians' and patients' perceptions of the use of artificial intelligence decision aids to inform shared decision making: A systematic review," *The Lancet*, vol. 398, p. S80, Nov. 2021.

27. Healthcare provider 2, "Personal communication on traditional (paper and pen) and electronic means for recording and maintaining patients' records," January 15 2025, kelantan.

28. Healthcare provider1, "Personal communication on traditional (paper and pen) and electronic means for recording and maintaining patients' records," January 14 2025, kelantan.

29. E. J. Hegedus and J. Mood, "Clinimetrics corner: The many faces of selection bias," *Journal of Manual & Manipulative Therapy*, vol. 18, no. 2, pp. 69–73, 2010. [Online]. Available: https://doi.org/10.1179/106698110X12640740712699

30. S. Johri, J. Jeong, and B. A. T. et al., "An evaluation framework for clinical use of large language models in patient interaction tasks," *Nature Medicine*, 2025. [Online]. Available: https://doi.org/10.1038/s41591-024-03328-5

31. I.-C. Jung, K. Schuler, M. Zerlik, S. Grummt, M. Sedlmayr, and B. Sedlmayr, "Overview Of Basic Design Recommendations For User-Centered Explanation Interfaces For AI-Based Clinical Decision Support Systems: A Scoping Review," *Digital Health*, vol. 11, 2025.

32. P. Kumar, S. Chauhan, and L. K. Awasthi, "Artificial intelligence in healthcare: review, ethics, trust challenges & future research directions," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105894, 2023.

33. Y. Kumar, A. Koul, R. Singla *et al.*, "Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 8459–8486, 2023.

34. H. Lakkaraju and O. Bastani, "The Hidden Cost of Opacity: A Study on Trust in Healthcare AI," *Nature Digital Medicine*, vol. 6, no. 1, pp. 45–58, 2023.

35. S. I. Lambert, M. Madi, S. Sopka *et al.*, "An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals," *NPJ Digital Medicine*, vol. 6, no. 1, p. 111, 2023.

36. J. D. Lee and K. A. See, "Trust In Automation: Designing For Appropriate Reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.

37. D. Long and B. Magerko, "What is AI literacy? competencies and design considerations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–16. [Online]. Available: https://doi.org/10.1145/3313831.3376727

38. Ministry of Home Affairs, "Sistem Pemprofilan Rekrutmen 2.0 (SPIER 2.0) Kementerian Dalam Negeri (English Translation: Recruitment Profiling System 2.0 (SPIER2.0) Of The Ministry Of Home Affairs)," 2022, retrieved 12 January 2025. [Online]. Available: http://psikom.moha.gov.my/spier_2

39. M. Mitchell, S. Wu, A. Zaldivar *et al.*, "Model Cards For Model Reporting: Building Trust Through Transparency," in *International Conference on AI Ethics*. Springer, 2022, pp. 325–331.

40. M. Mooghali, A. M. Stroud, D. W. Yoo *et al.*, "Trustworthy and ethical ai-enabled cardiovascular care: A rapid review," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 247, 2024. [Online]. Available: https://doi.org/10.1186/s12911-024-02653-6

41. S. Nundy and T. Montgomery, "The Path To Clinical AI Adoption: Trust As The Cornerstone," *JAMA*, vol. 329, no. 8, pp. 631–632, 2023, this paper demonstrates that trust levels directly correlate with adoption rates (r=0.78) and found that departments with structured AI training programs had 3x higher adoption rates.

42. R. Pawson and N. Tilley, *Realistic Evaluation*. Thousand Oaks, CA: Sage Publications, 1997.

43. A. Perivolaris, C. Adams-McGavin, Y. Madan *et al.*, "Quality of interaction between clinicians and artificial intelligence systems. a systematic review,"

Future Healthcare Journal, vol. 11, no. 3, p. 100172, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2514664524015625

44. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1135–1144.

45. E. M. Rogers, *Diffusion Of Innovations*, 5th ed. Free Press, 2003.

46. M. G. Seneviratne, R. C. Li, M. Schreier *et al.*, "User-centred design for machine learning in health care: A case study from care management," *BMJ Health & Care Informatics*, vol. 29, p. e100656, 2022.

47. E. Sezgin, "Artificial intelligence in healthcare: Complementing, not replacing, doctors and healthcare providers," *Digital Health*, vol. 9, p. 20552076231186520, 2023.

48. J. Shen, C. J. P. Zhang, B. Jiang *et al.*, "Artificial intelligence versus clinicians in disease diagnosis: Systematic review," *JMIR Med Inform*, vol. 7, no. 3, p. e10010, Aug. 2019. [Online]. Available: http://medinform.jmir.org/2019/3/e10010/

49. A. Smith, N. Thomas, C. Snoswell *et al.*, "Telehealth for global emergencies: Implications for coronavirus disease 2019 (covid-19)," *Journal of Telemedicine and Telecare*, vol. 26, no. 5, pp. 309–313, 2020.

50. E. Steerling, E. Siira, P. Nilsen, P. Svedberg, and J. Nygren, "Implementing AI in healthcare-the relevance of trust: A scoping review," *Frontiers in Health Services*, vol. 3, p. 1211150, Aug 2023.

51. T. Toscos, M. Drouin, J. Pater, M. Flanagan, R. Pfafman, and M. J. Mirro, "Selection biases in technology-based intervention research: Patients' technology use relates to both demographic and health-related inequities," *J Am Med Inform Assoc*, vol. 26, no. 8-9, pp. 835–839, 2019.

52. C. M. van Leersum and C. Maathuis, "Human centred explainable ai decision-making in healthcare," *Journal of Responsible Technology*, vol. 21, p. 100108, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666659625000046

53. V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003.

54. V. Vo, G. Chen, Y. S. J. Aquino, S. M. Carter, Q. N. Do, and M. E. Woode, "Multi-stakeholder preferences for the use of artificial intelligence in healthcare: A systematic review and thematic analysis," *Social Science & Medicine*, vol. 338, p. 116357, 2023.

55. R. Walker, "Co-creating support strategies for cancer survivorship," Jan. 2025, video. [Online]. Available: https://designthinkingforhealth.org/

56. C. Warzel, "Ai has become a technology of faith," 2024, accessed: 2024-12-25. [Online]. Available: https://www.theatlantic.com/technology/archive/2024/07/thrive-ai-health-huffington-altman-faith/678984/?utm_source=chatgpt.com

57. World Health Organization, *International Classification Of Diseases (ICD-11)*, World Health Organization, Geneva, Switzerland, 2019, https://icd.who.int/.

58. M. Yin, J. Wortman Vaughan, and H. Wallach, "Understanding The Effect of Accuracy on Trust in Machine Learning Models," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, pp. 1–12.

59. H. A. Younis, T. A. E. Eisa, M. Nasser *et al.*, "A systematic review and meta-analysis of artificial intelligence tools in medicine and healthcare: Applications, considerations, limitations, motivation and challenges," *Diagnostics*, vol. 14, no. 1, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/14/1/109

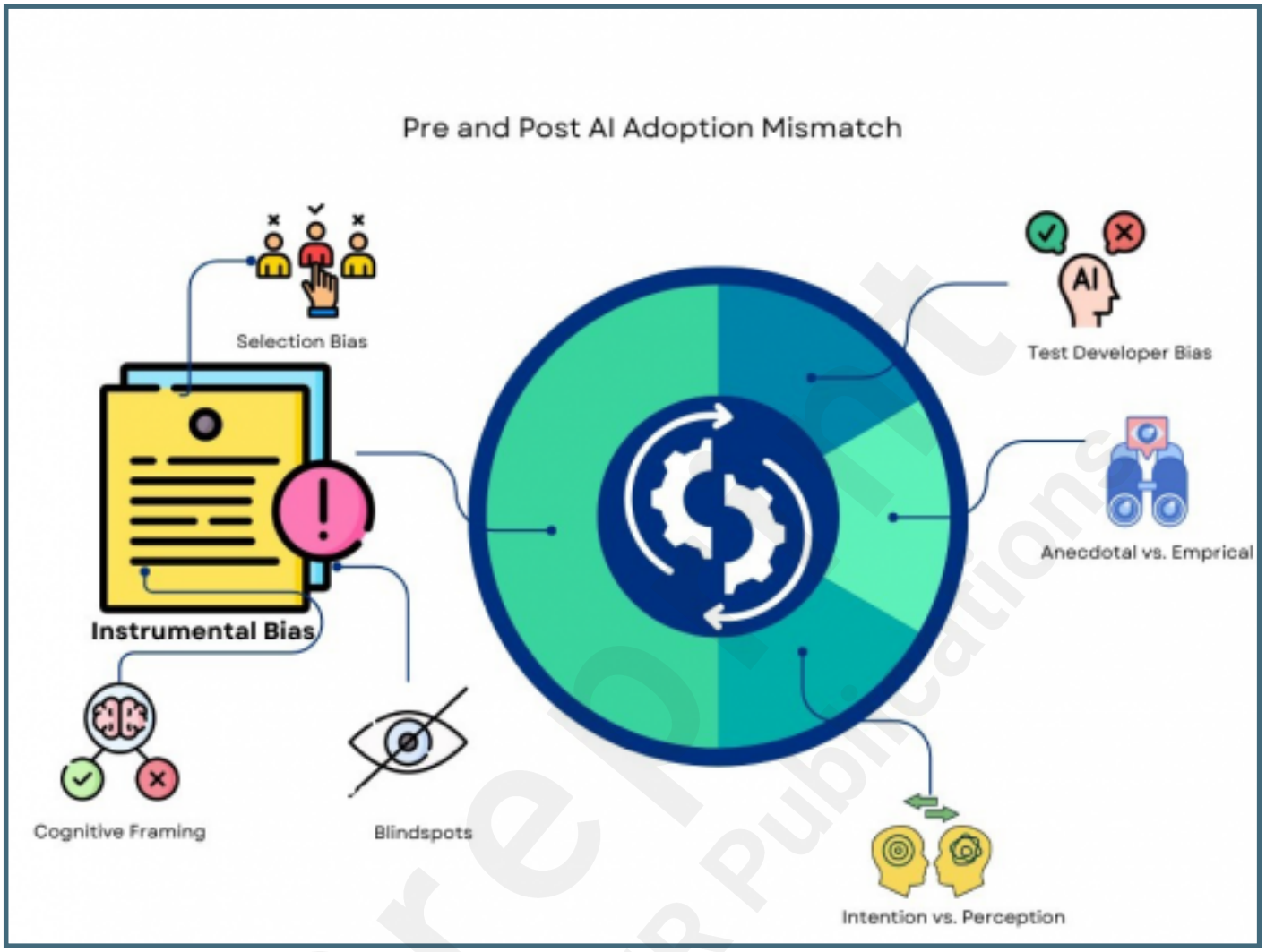
60. Y. Zhang, Q. V. Liao, R. K. E. Bellamy, and K. R. Varshney, "The Role Of Transparency In Building Clinical Trust," *JAMA Digital Health*, vol. 5, no. 2, p.

e230145, 2023.

Supplementary Files

Figures

Potential Reasons for the Gaps between Pre and Post AI Adoption Outcomes.



The five stages of iterative, non-linear Design Thinking.